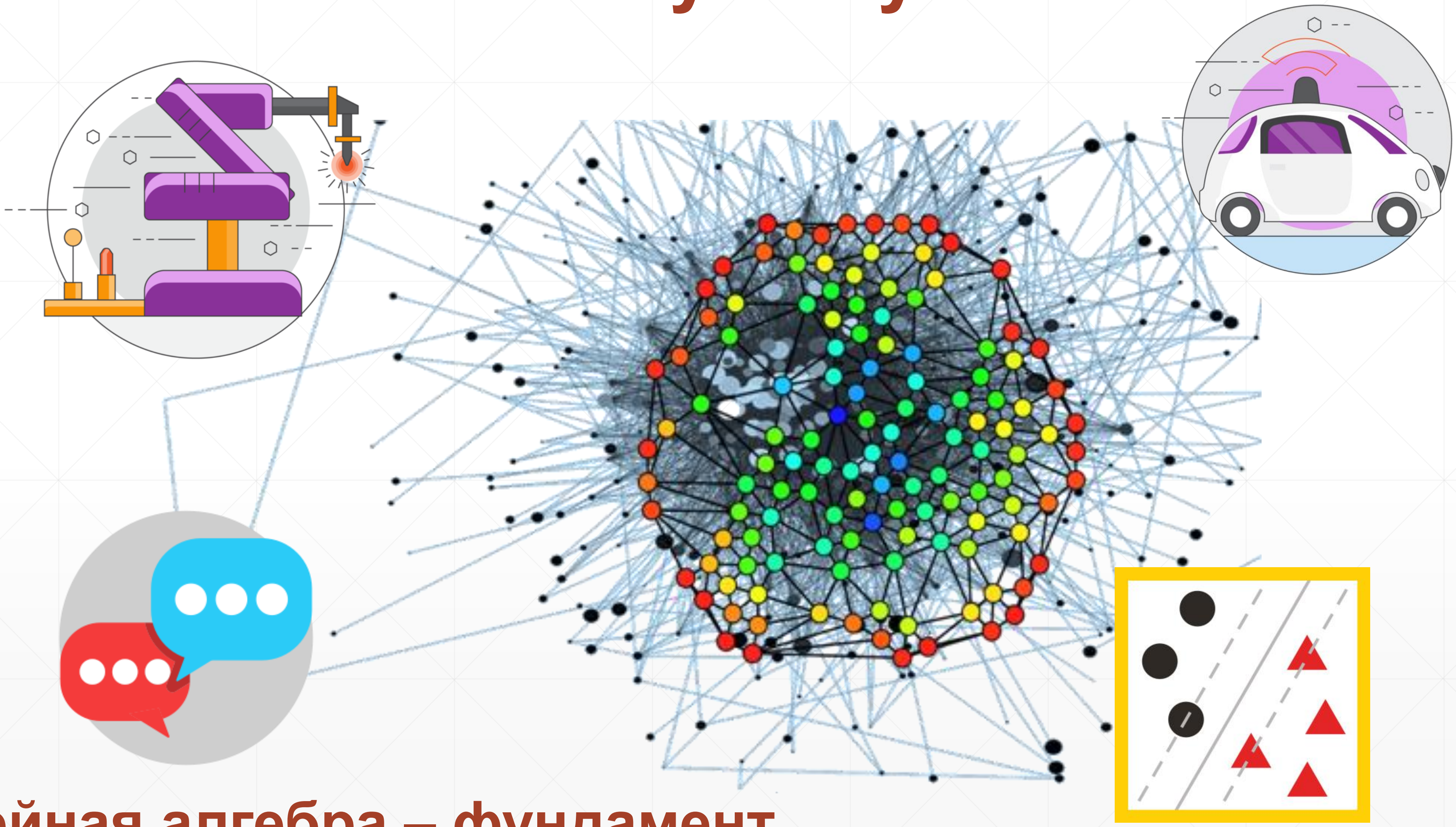
Основы Линейной Алгебры

Гончаров Павел kaliostrogoblin3@gmail.com
Нестереня Игорь nesterione@gmail.com

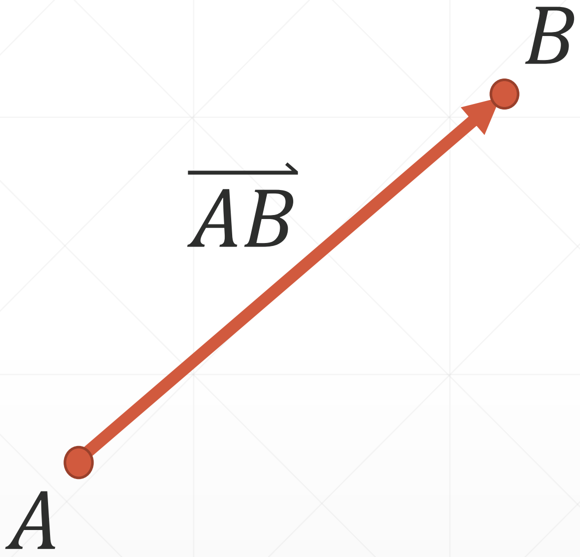
Понимание – ключ к успеху!



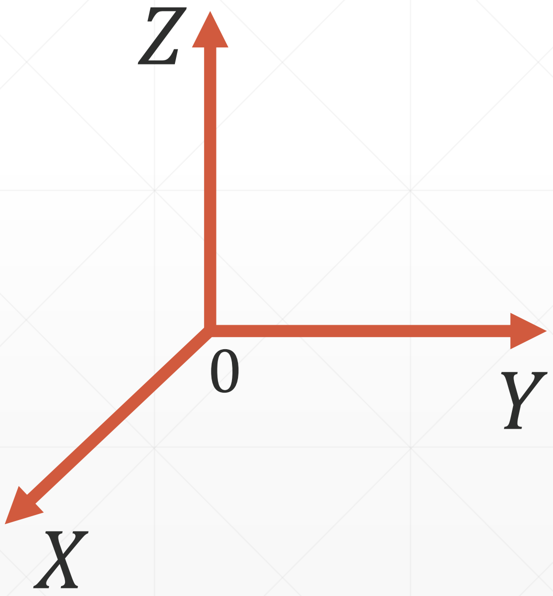
Линейная алгебра – фундамент машинного обучения

Что мы знаем со школы?

Вектор



Система координат



Матрица

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Основные понятия. Скаляр

Скаляр – это величина, определяемая одним числом, значение которого остается неизменным даже при изменении пространственной системы координат.

Примеры скаляров:

- длина;
- время;
- угол;
- **скорость обучения**.

Когда мы определяем скаляр, мы указываем множество значений, к которым принадлежит скаляр:

$s \in \mathbb{R}$ – скаляр s может быть любым вещественным числом.

$n \in \mathbb{N}$ – переменная принадлежит множеству натуральных чисел

Основные понятия. Вектор

Вектор – это массив значений, который записывают в квадратных скобках. Вектор представляет собой матрицу размером $n \times 1$.

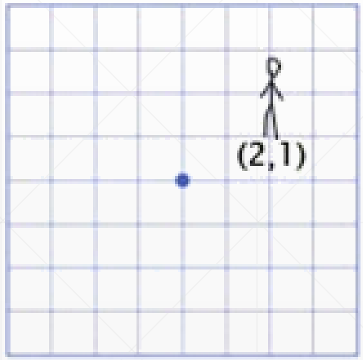
$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \text{ где } n \text{ – это длина вектора. } x \in \mathbb{R}^n$$

Индексация:

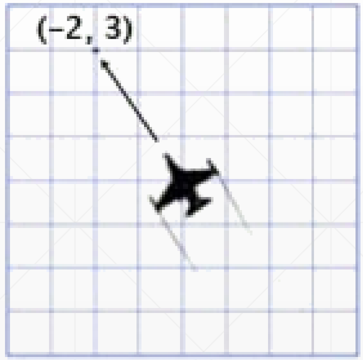
$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad x_2 = 2, \quad x_{\{1,3\}} = (1,3)$$

Примеры:

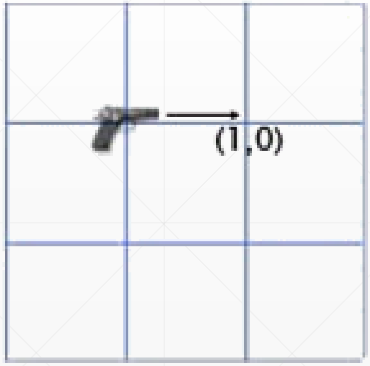
В ИГРАХ:



Местоположение



Скорость



Направление

В МАШИННОМ ОБУЧЕНИИ:

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

Основные понятия. Матрица

Матрица – это двумерный массив значений, т.е. каждый элемент <u>имеет два индекса</u> – индекс строки и индекс столбца. Про матрицу говорят, что она размерности $n \times m$, где n – это число строк, а m – число столбцов.

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}, X \in \mathbb{R}^{n \times m}$$

Матрица признаков

$$X = \begin{bmatrix} 5.1 & 3.5 & 1.4 & 0.2 \\ 4.9 & 3.0 & 1.4 & 0.2 \end{bmatrix}$$

Строка – это набор признаков для одного экземпляра из базы.

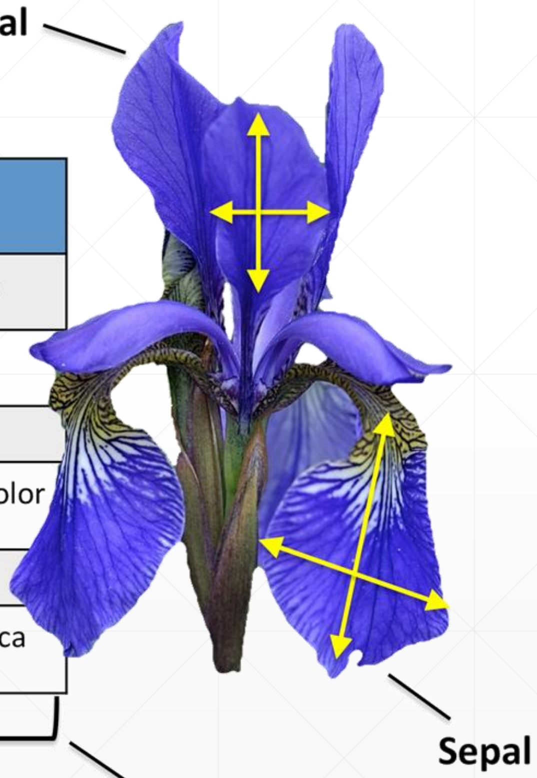
Столбец – значения определенного признака для всех примеров.

Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

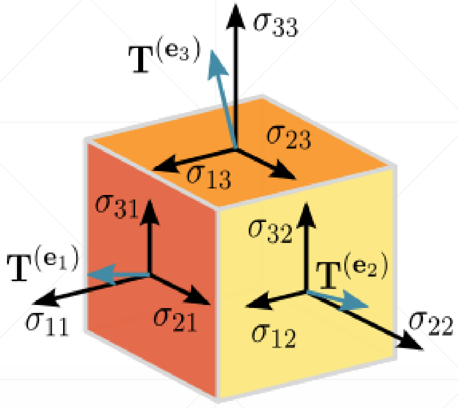
Features
(attributes, measurements, dimensions)

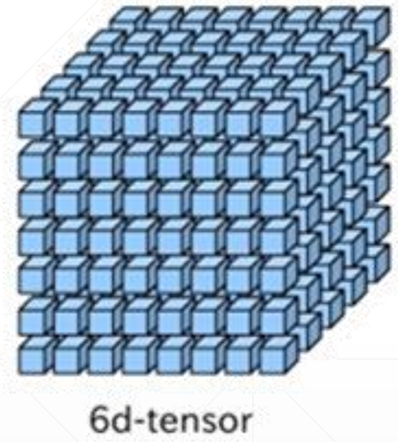
Class labels
(targets)

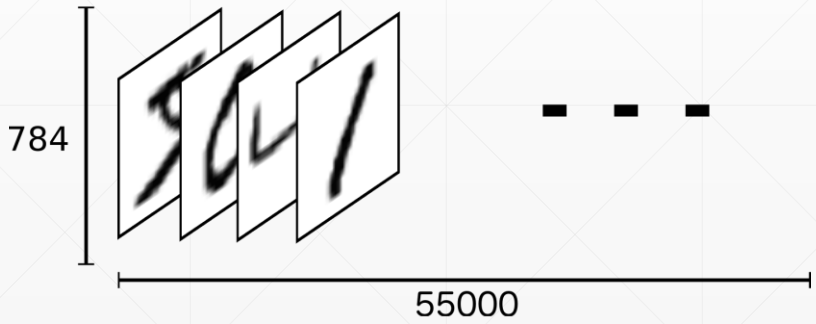


Основные понятия. Тензор

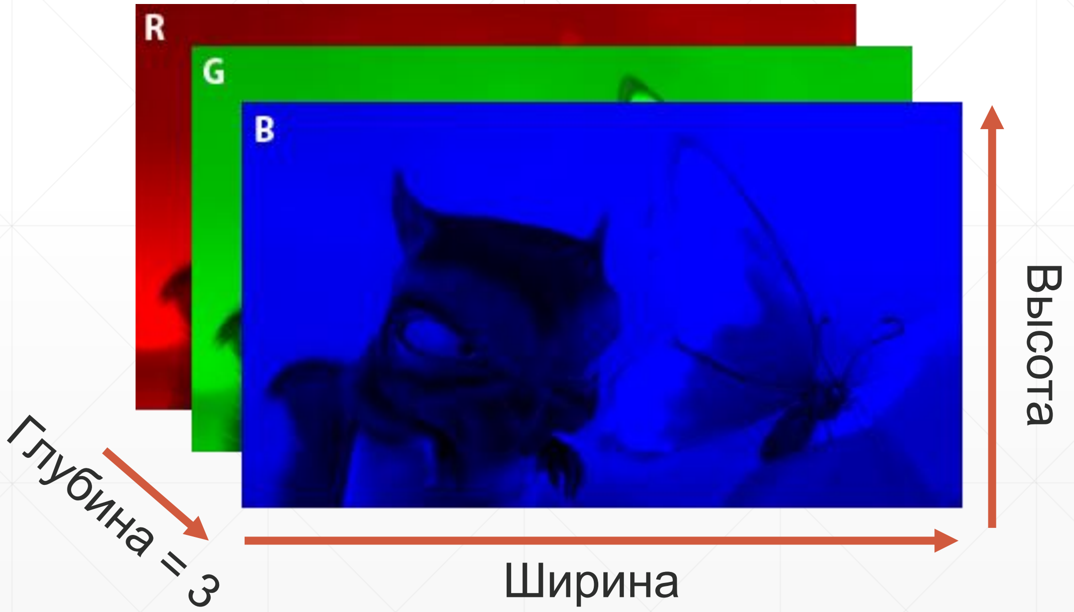
Тензор – это массив значений, расположенных на регулярной сетке с переменным числом осей. Тензор – это в, какой-то степени, многомерная матрица.







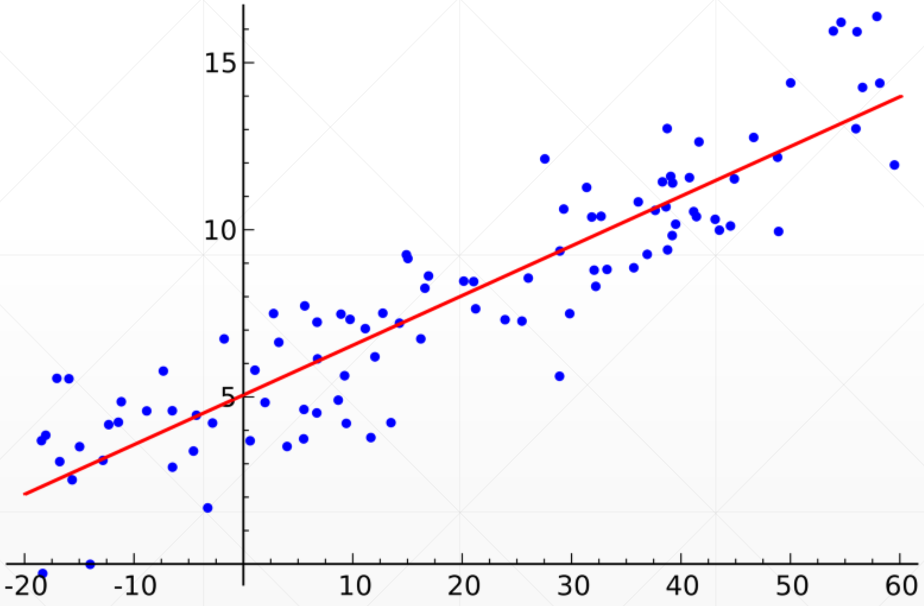
Любое RGB изображение – это тензор



Машинное обучение. Модель

Модель в машинном обучении – это некая гипотеза $h_\theta(x)$, согласно которой мы делаем предсказание, классификацию, разделение данных и т.д.

Пример модели: **линейная регрессия**



Модель линейной регрессии:

Параметры модели, веса

$$\boxed{h_\theta(x)} = \boxed{W}x + \boxed{b}$$

Гипотеза h с параметрами θ, зависящая от x

Набор признаков

Почему <u>линейная</u> регрессия?

$y = kx + b$ – это линейная функция

<u>**Основная задача**</u> машинного обучения– при тренировке модели подобрать такие параметры, при которых <u>*ошибка минимальна*</u>!

Линейная алгебра. Транспонирование

Транспонирование – операция, в результате которой все столбцы становятся строками, а все строки – столбцами.

$$\begin{bmatrix} 6 & 4 & 24 \\ 1 & -9 & 8 \end{bmatrix}^{T} = \begin{bmatrix} 6 & 1 \\ 4 & -9 \\ 24 & 8 \end{bmatrix}$$

Вектор очень удобно записывать в такой форме:

$$x = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

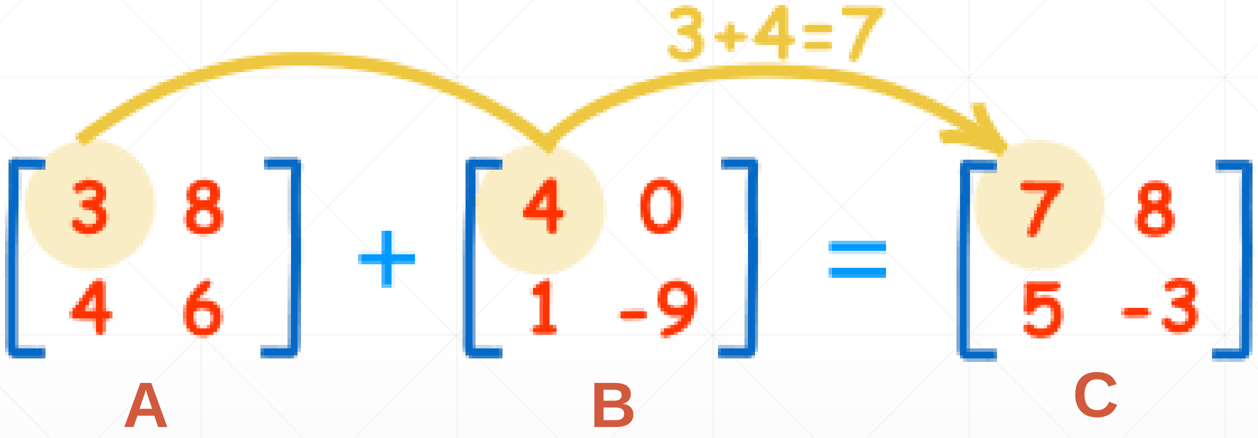
Для скаляра операция транспонирования <u>ничего не изменяет</u>, т.е. если a – это скаляр, то $a = a^T$

В случае с тензором нужно указать, <u>какие размерности поменять местами</u>. Например, в библиотеке *tensorflow* есть функция *tf.transpose*, содержащая аргумент *perm*, в который нужно передать список из размерностей. Пример:

tf.transpose(x, perm=[0,2,1]*)* => в этом случае мы меняем столбцы тензора x на срезы по глубине.

Линейная алгебра. Сложение

Сложение – операция, аналогичная обычному сложению чисел. Выполняется над матрицами, имеющими одинаковую размерность: $\cancel{X_{n\times m} + Y_{n\times(m+1)}}$. Чтобы вычислить сумму двух матриц или векторов, нужно сложить соответствующие элементы:



Математическая запись:

$$A_{2\times2} + B_{2\times2} = C_{2\times2}$$
$$A_{0,0} + B_{0,0} = C_{0,0} =$$
$$= 3 + 4 = 7$$

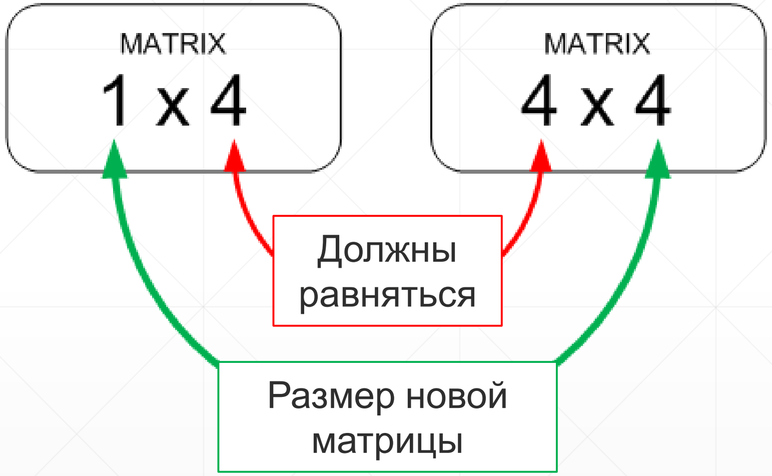
Также мы можем прибавить скаляр к матрице, путем сложения скаляра с каждым элементом матрицы. Можно складывать вектор с матрицей, прибавляя вектор к каждому столбцу матрицы **(broadcasting)**: $C_{i,j} = A_{i,j} + b_j$

Вычитание – это операция сложения со знаком минус.

Линейная алгебра. Умножение

$$AB \neq BA$$

Матричное умножение. Результатом произведения матрицы размерностью $n \times m$ на матрицу $m \times l$ будет новая матрица размером $n \times l$.
<u>Число столбцов первого множителя должно равняться числу строк второго</u>.



$$C_{i,j} = \sum_k A_{i,k} B_{k,j}$$

Матричное умножение «Dot product»

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \times \begin{bmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{bmatrix} = \begin{bmatrix} 58 & 64 \\ & \end{bmatrix}$$

$$\mathbf{1 \times 8 + 2 \times 10 + 3 \times 12 = 64}$$

2x4=8

$$2 \times \begin{bmatrix} 4 & 0 \\ 1 & -9 \end{bmatrix} = \begin{bmatrix} 8 & 0 \\ 2 & -18 \end{bmatrix}$$

Произведение матрицы и скаляра

Линейная алгебра. Уравнение модели

Пример: у нас есть гипотеза $h_\theta(x)$, благодаря которой можно предсказать стоимость дома, зная его площадь, в футах квадратных (1 фут$^2 \approx 0.0929$ м^2).

Площадь домов:

$$\text{x} = \begin{bmatrix} 2104 \\ 1416 \\ 1534 \\ 852 \end{bmatrix}$$, где x_1 – размер первого дома

Гипотеза:

$$h_\theta(x) = -40 + 0.25x$$

Стоимость домов согласно нашей гипотезе

$$\underset{4\times 2}{\begin{bmatrix} 1 & 2104 \\ 1 & 1416 \\ 1 & 1534 \\ 1 & 852 \end{bmatrix}} \times \underset{2\times 1}{\begin{bmatrix} -40 \\ 0.25 \end{bmatrix}} = \underset{4\times 1}{\begin{bmatrix} -40\times 1 + 0.25\times 2104 \\ -40\times 1 + 0.25\times 1416 \\ -40\times 1 + 0.25\times 1534 \\ -40\times 1 + 0.25\times 852 \end{bmatrix}} = \begin{bmatrix} 486 \\ 314 \\ 343.5 \\ 173 \end{bmatrix}$$

Матрица признаков

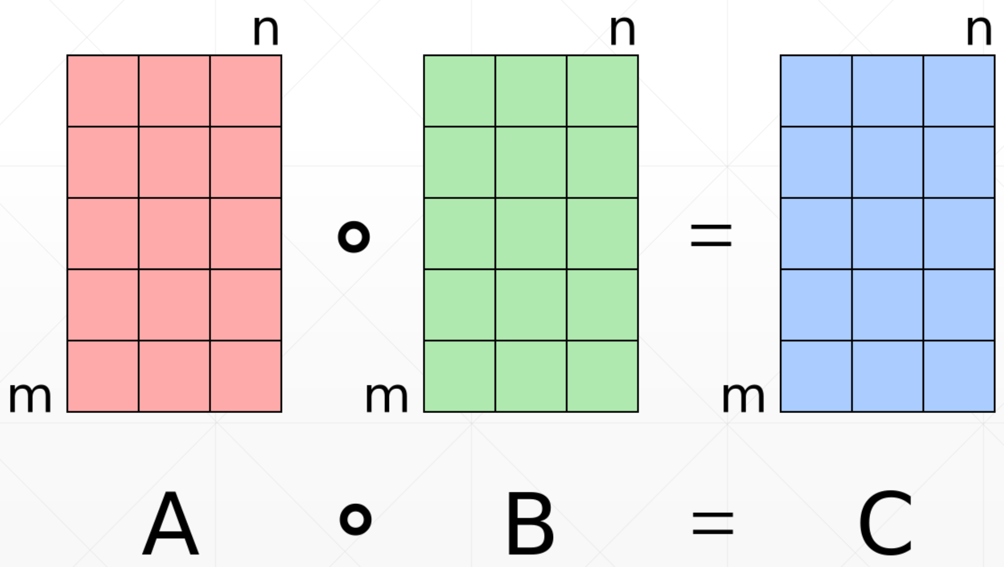
Параметры модели

Вектор предсказаний

Линейная алгебра. Произведение Адамара

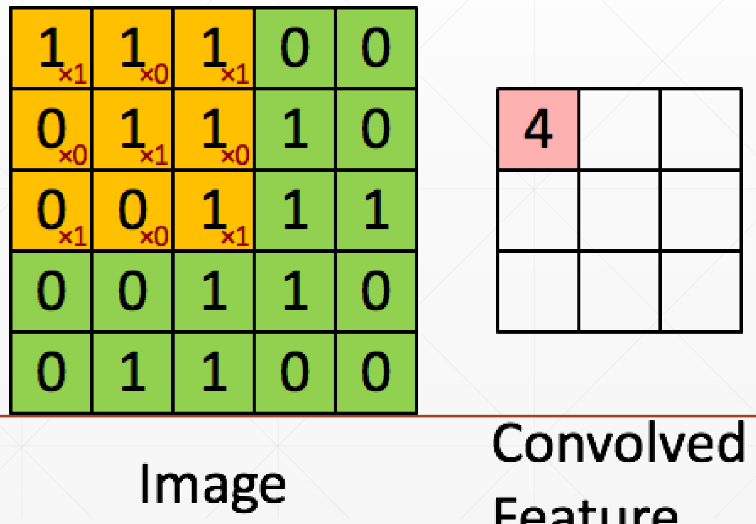
Произведение Адамара или покомпонентное произведение – операция над двумя матрицами/векторами одинаковой размерности, в результате которой создается новая матрица, где каждый элемент i,j это произведение элементов i,j исходных матриц.

Математическая запись: $A \circ B = C$



Произведение Адамара не определено для матриц разной размерности.

<u>Пример использования</u> – сверточная сеть. Выполнение **операции свертки** является покомпонентным произведением.



Частные случаи. Диагональная, нулевая матрицы

Нулевой называется матрица, все элементы которой равны нулю. Если умножить нулевую матрицу на другую матрицу правильной размерности, то результатом станет нулевая матрица.

Диагональная матрица – это матрица, в которой все элементы, кроме тех, что находятся на главной диагонали, равны нулю. Частный случай диагональной матрицы – это единичная матрица.

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Если умножить матрицу верной размерности на единичную матрицу, то получится исходная матрица:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \qquad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad AI = A$$

$$\begin{bmatrix} 1 \times 1 + 2 \times 0 & 1 \times 0 + 2 \times 1 \\ 3 \times 1 + 4 \times 0 & 3 \times 0 + 4 \times 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

Частные случаи. Обратная матрица

Обратная матрица – это матрица A^{-1}, при умножении на которую исходная матрица A дает единичную матрицу I:

$$AA^{-1} = I$$

<u>Не у всех матриц есть обратная!</u> Обратная матрица может быть только у **квадратной** матрицы с **определителем не равным нулю.**

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\underbrace{ad-bc}_{\text{determinant}}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Норма. L2 и L1 нормы, Фробениуса

Иногда нужно как-то измерить размер вектора (не длину). В машинном обучении для этих целей используется **функция нормы**. Обозначается норма, как L^p:

$$\|x\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}, \text{для } p \in \mathbb{R}, p \geq 1$$

Норма в простом понимании – это расстояние от нулевой точки до x. L2 норма, также известная как Евклидовая норма, и вычисляется, как: $\|x\|_2 = x^T x$.
В машинном обучении часто нужно различать, какие элементы равны нулю, а какие нет. Для того, чтобы увеличить разницу, используется L1 норма:

$$\|x\|_1 = \sum_i |x_i|$$

Норма Фробениуса – это частный случай L2 нормы для матриц: $\sqrt{\sum_{i,j} A_{i,j}^2}$

Пора создать свою матрицу!